OMCHOCHO README FILE

July 26 2016

OMI CHOCHO Column (JJA 2005)

10^{14} molecules cm^{-2}

< 0.00   2.00   4.00   6.00   8.00 >
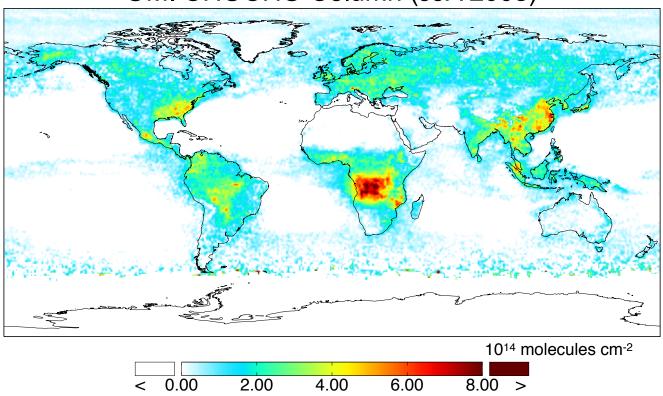
1 Overview

This document provides a brief description of the OMCHOCHO data product. OMCHOCHO contains total column CHOCHO and ancillary information retrieved from OMI global and spatial zoom mode measurements using a retrieval algorithm that is based on nonlinear least-squares fitting originally developed for GOME, and adapted for the OMI instrument. In global mode, i.e., global coverage in one day, each file contains a single orbit of data covering a swath approximately 2,600 km wide from pole to pole (sunlit portions only). Fitting uncertainties for the CHOCHO slant columns (single measurement) typically range between 100-400%, with the lower end of this range over CHOCHO hotspots. Uncertainties in the air mass factor (AMF), used to convert slant to vertical columns, are estimated to be 30%. Hence the total uncertainties of the CHOCHO vertical columns typically range within 104-401%. The single measurement errors are dominated by random noise, and can be reduced by averaging. The image above shows monthly average CHOCHO for June-August 2005, including only pixels with cloud fraction up to 20%.

2 Algorithm Description

The algorithm is based on the direct fitting of radiances and irradiances. In particular, and differing from what is commonly referred to as Differential Optical Absorption Spectroscopy (DOAS) fitting, radiances are not divided by irradiances, no logarithms are taken of the spectra, and no high-pass filtering is applied.

The three main stages of the algorithm are (1) Radiance wavelength calibration, which finds the optimum wavelength registration for a representative swath of radiance measurements (usually in the middle of the orbit) and determines a common wavelength grid for auxiliary data bases (molecular reference cross sections, etc.); (2) On-line computation of a residual "common mode" spectrum calculated from the fitting residuals of the central portion of the orbit to account for systematic features not considered in the semi-empirical model; and (3) Nonlinear, least-squares fitting of all swath lines in the OMI granule. In each stage, the calibration or fitting is performed individually for the 60 cross-track pixels of an OMI swath line. For improved numerical stability, radiances and irradiances are divided by their respective averages over the fitting window; in other words, they are normalized to values ~ 1.

CHOCHO fitting is performed in the spectral window 435 - 461 nm, within the visible channel of the OMI instrument. The model that is fitted to the measurements consists of a solar irradiance reference, attenuated by contributions from CHOCHO (the target gas), inelastic (rotational Raman, or Ring) scattering, and interferences from other atmospheric gases, including O_3, NO_2, H_2O) and the O_2-O_2 collision complex, and a fixed absorption due to liquid water derived from a spectral fit to 385 - 470 nm. The model also includes additive and multiplicative closure polynomials and parameters for spectral shift and, potentially, squeeze, as well as an undersampling correction (Chance et al., 2005) and a common mode spectrum, both of which are computed on-line. The common mode spectra (one per cross-track position) are the average of several hundred fitting residuals. They include any instrument effects that are unrelated to molecular scattering and absorption cross sections. The least-squares fit is mostly unconstrained, with the exception of selected parameters, including the spectral shift, which are constrained in order to prevent problems arising from out-of-bounds values.

The results from the spectral fitting are CHOCHO slant columns. These are converted to vertical columns using a look-up table of AMFs. AMF tables have been pre-computed with a radiative transfer model using climatological CHOCHO profiles. They include the effect of clouds (non-scattering, reflecting cloud top) and surface albedo. The AMFs used for the conversion from slant to vertical columns are provided in the data product for all ground pixels. For global mode granules, the CHOCHO retrieval uses cloud fraction and cloud pressure from the OMI O_2-O_2 cloud product, OMCLDO2 and the OMI Lambertian Equivalent Reflectance climatology (OMLER, Kleipool et al. (2008)) for the surface reflectance. Scattering weights and the CHOCHO climatology used in the AMF calculation are provided in the level 2 files fields GasProfile and ScatteringWeights. The altitude pressure coordinate for these two fields is provided separately in the ClimatologyLevels field.

The algorithm uses a reference sector correction method to reduce cross-track striping of the CHOCHO columns. The reference sector is over the Sahara Desert, where CHOCHO concentrations are expected to be negligable. The uncorrected columns are reported in the field ColumnAmount, and the corrected columns in the field ColumnAmountDestriped. It is strongly recommended that the ColumnAmountDestriped field be used for analysis. More details of the algorithm can be found in Chan Miller et al. (2014).

3 Data Quality Assessment

Pixels affected by the row anomaly have been flagged in the Geolocation Field XtrackQualityFlags and XtrackQualityFlagsExpanded. Their use is discouraged. The CHOCHO data product provides RMS (data field FittingRMS) and one standard deviation (1σ) fitting uncertainties (ColumnUncertainty), as derived from the fitting covariance matrix. The uncertainties do not include contributions from uncertainties in the reference cross sections. The main guidance to data quality provided with the CHOCHO columns is

Table 1: MainDataQualityFlag value descriptions

Value	Classification	Rationale
0	Good	Column value present and passes all quality checks; data may be used with confidence
1	Suspect	Caution advised because one or more of the following conditions are present: • FitConvergenceFlag is < 300 (but > 0): convergence at noise level • Column+2σ uncertainty < 0 (but Column+3σ uncertainty ≥ 0) • Absolute column value $>$ MaximumColumnAmount (1×10^{16} molec cm^{-2})
3	Bad	Avoid using data because one or more of the following conditions are present: • FitConvergenceFlag is < 0: abnormal termination, no convergence • Column+3σ uncertainty < 0
≤ -1	Missing	No column values have been computed; entries are missing

the MainDataQualityFlag, which is set to any of four values (0, 1, 2, and $\leq$ -1) based on the outcome of the fitting process. This flag should be used for data screening prior to use of each individual OMI pixel column. Additional information on the convergence of the fit is provided in a fitting diagnostic flag (FitConvergenceFlag); this flag should be consulted if more detailed information on the fitting process is desired.

4 Cloud Information

The OMCHOCHO product file contains two data fields with cloud information, taken from the OMCLDO2 data product and provided in slightly modified form: (1) AMFCloudFraction contains the cloud fractional cover (between 0 and 1) in the OMI pixel, as it has been used in the AMF calculation; it has been adjusted from the original OMCLDO2 data field, first by forcing all non-missing values into the interval [0,1], and second by replacing any missing values of pixels for which CHOCHO retrievals were successful by values from the ISCCP D2 cloud fraction climatology. (2) AMFCloudPressure contains the cloud pressure (in hPa) with any missing values for pixels of successful CHOCHO retrievals replaced by ISCCP D2 cloud pressure climatological values. As the names of the data fields suggest, they have been used in the calculation of the CHOCHO air mass factor. It must be emphasized that the information on cloud parameters included with the OMCHOCHO product is not identical to the original OMCLDO2 product, and that it may not be used for independent cloud studies. The data fields are mainly provided to facilitate cloud screening of the CHOCHO columns.

5 Product Description

A 2600 km wide OMI swath contains 60 cross-track pixels, ranging in size from 14×24 km^2 (along $\times$ across track) in the center of the swath to about 28×150 km^2 at the edges of the swath (median: 15×33 km^2). The pixels on the swath are not symmetrically aligned on the line perpendicular to the orbital

plane. However, the latitude and longitude provided with each pixel represents the location of each pixel on the ground to a fraction of a pixel. The OMCHOCHO product is written as an HDF-EOS5 swath file. A single OMCHOCHO file contains information retrieved from each OMI pixel over the sunlit portion of the orbit (a.k.a. an OMI granule). The information provided in these files include: Geodetic longitude and latitude, solar and line-of-sight zenith and azimuth angles, total column CHOCHO with RMS and 1σ fitting uncertainties, longitude and latitude corner coordinates for each OMI pixel, and a range of ancillary parameters that provide information to assess data quality. Average values over an OMI granule for the HCHO total column, uncertainties, and RMS, as well as the percent values of "good" (converged and columns positive within 2σ fitting uncertainties) and "bad" (failed convergence or truly negative columns) provide general, granule-based information on data quality. For questions and comments related to the OMCHOCHO dataset please contact Christopher Chan Miller. Please send a copy of your e-mail to Kelly Chance, who has the overall responsibility for this product.

6   Selected List of Elements in an OMCHOCHO Output File

The tables below show a selected list of data elements in an OMCHOCHO HDF-EOS5 output file. The tables are divided into (a) Swath Dimensions, (b) Geolocation Fields, and (c) Data Fields. The selection of the listed Geolocation and Data Fields is entirely arbitrary and made solely to facilitate the identification of what is assumed will be the most-used parameters from the OMCHOCHO data product. No such distinction is made in the HDF-EOS5 product file itself. Naturally, whether or not any part of the product is of interest ultimately depends on the application.

Table 2: Swath Dimensions

Field Name	Field Type	Description
nTimes	HE5T_NATIVE_INT	Number of swath lines in an OMI granule (usually about 1650)
nXtrack	HE5T_NATIVE_INT	Number of cross-track positions in a swath line (usually 30 or 60)
nUTCdim	HE5T_NATIVE_INT	Number of elements in a single TimeUTC field entry (6)
Nlevels	HE5T_NATIVE_INT	Number of GEOS-Chem cliimatology levels

7   Recommendations for gridding L2 data

The retrieved column densities are a function of the CHOCHO concentration vertical profile. The archived AMF values have been computed using a CHOCHO climatology from the GEOS-Chem chemical transport model. To quantitatively compare simulated CHOCHO columns to the OMI data, it is necessary to recompute the airmass factors (A) using CHOCHO concentration profiles from the new simulation.

To use the scattering weights provided, profiles from the new simulation must be interpolated to the scattering weight vertical grid. Let $\boldsymbol{P^{low}}(\equiv$'ClimatologyAltitude') be a vector of pressures (in hPa) at the bottom of each vertical level, and lmx be the number of vertical levels. The vector of midpoint pressures ($\boldsymbol{p^{mid}}$) is

$$\boldsymbol{p^{mid}} = 0.5 \times \left[\boldsymbol{p^{low}_{1:lmx}} + (\boldsymbol{p^{low}_{2:lmx}}, 0.01)\right] \tag{1}$$

Similarly the vector of pressure differences over each layer (Δp) is

$$\Delta \boldsymbol{p} = \boldsymbol{p^{low}_{1:lmx}} - (\boldsymbol{p^{low}_{2:lmx}}, 0.01) \tag{2}$$

Table 3: Geolocation Fields of prime interest

Field Name	Field Type	Dimensions	Description
Latitude	HE5T_NATIVE_FLOAT	nXtrack,nTimes	Geodetic latitude [deg] at the center of the ground pixel
Longitude	HE5T_NATIVE_FLOAT	nXtrack,nTimes	Geodetic longitude [deg] at the center of the ground pixel
SolarZenithAngle	HE5T_NATIVE_FLOAT	nXtrack,nTimes	The solar zenith angle [deg] at the center of the ground pixel
ViewingZenithAngle	HE5T_NATIVE_FLOAT	nXtrack,nTimes	The viewing zenith angle [deg] at the center of the ground pixel
TimeUTC	HE5T_NATIVE_INT16	nUTCdim,nTimes	UTC value of the TAI93 time. The 6 different elements of the UTC string YYYY-MM-DD hh:mm are stored in the 6 arrays positions
XtrackQualityFlags	HE5T_NATIVE_INT8	nXtrack,nTimes	Cross-Track quality flags as set in the L1b to flag row anomaly
XtrackQualityFlagsExpanded	HE5T_NATIVE_INT16	nXtrack,nTimes	Cross-Track quality flags as set in the L1b to flag row anomaly. Expanded human-readable version of XtrackQualityFlags

Table 4: Data Fields of prime interest

Field Name	Field Type	Dimensions	Description
AirMassFactor	HE5T_NATIVE_DOUBLE	nXtrack,nTimes	Molecule specific air mass factor for each ground pixel, including scattering weights, clouds, and vertical distribution of CHOCHO
AMFCloudFraction	HE5T_NATIVE_FLOAT	nXtrack,nTimes	Cloud fraction from external cloud ESDT, adjusted such that it falls within the range of [0, 1]
AMFCloudPressure	HE5T_NATIVE_FLOAT	nXtrack,nTimes	Cloud pressure from external cloud ESDT, internally adjusted such that it falls within the range AMF table cloud pressure variable range of [0, 900]
Albedo	HE5T_NATIVE_DOUBLE	nXtrack,nTimes	Ground pixel albedo from the OMLER database for the central wavelength of the fitting window
ColumnAmount	HE5T_NATIVE_DOUBLE	nXtrack,nTimes	Total column amount [mol/cm2] for each ground pixel
ColumnAmountDestriped	HE5T_NATIVE_DOUBLE	nXtrack,nTimes	Total column amount [mol/cm2] for each ground pixel with stripe offset correction (**recommended**)
ColumnUncertainty	HE5T_NATIVE_DOUBLE	nXtrack,nTimes	Total column amount uncertainty [mol/cm2] for each ground pixel
GasProfile	HE5T_NATIVE_DOUBLE	nXtrack,nTimes, nLevels	Gas profiles used in the AMFs calculation
MainDataQualityFlag	HE5T_NATIVE_INT16	nXtrack,nTimes	Main flag to indicate data quality (see above)
PixelCornerLatitudes	HE5T_NATIVE_FLOAT	nXtrack+1, nTimes+1	The geodetic latitudes [deg] of the corner coordinates of the OMI ground pixels.
PixelCornerLongitudes	HE5T_NATIVE_FLOAT	nXtrack+1, nTimes+1	The geodetic longitudes [deg] of the corner coordinates of the OMI ground pixels.
ScatteringWeights	HE5T_NATIVE_DOUBLE	nXtrack,nTimes, nLevels	Scattering weights used in the AMFs calculation
ClimatologyAltitude	HE5T_NATIVE_DOUBLE	nXtrack,nTimes, nLevels	Pressure [hPa] at the bottom of each vertical level used in AMF calculation. The pressure at the top of the highest level is 0.01 hPa for all profiles.

Let $\boldsymbol{x}$ be the vector containing the modeled CHOCHO profile interpolated to $\boldsymbol{p^{mid}}$, and $\boldsymbol{S}$ ($\equiv$ 'ScatteringWeights') the vector of scattering weights. The airmass factor (A) using the new model profile is

$$A = \frac{\sum_{l=1}^{lmx} S_l x_l \Delta P_l}{\sum_{l=1}^{lmx} x_l \Delta P_l} \tag{3}$$

Let $A_0(\equiv$'AirMassFactor') and $\Omega_0(\equiv$'ColumnAmountDestriped') be the original air mass factor and CHOCHO vertical column amount respectively. The CHOCHO column amount (Ω) for the new profile is

$$\Omega = \Omega_0 \frac{A_0}{A} \tag{4}$$

Similarly the column amount uncertainty (σ) is given by

$$\sigma = \sigma_0 \frac{A_0}{A} \tag{5}$$

where $\sigma_0(\equiv$'ColumnUncertainty') is the original vertical column uncertainty.

To grid the L2 data, we recommend weighting the OMI observations by their uncertainties. For a given gridbox containing n CHOCHO column observations ($\Omega_1, \Omega_2, ..., \Omega_n$), the grid box weighted mean ($\overline{\Omega}$) can be computed as:

$$\overline{\Omega} = \sum_{i=1}^{n} w_i \Omega_i \tag{6}$$

w_i are the normalized weights. To weight by the column uncertainties (σ_i) the following expression for w_i can be used

$$w_i = \frac{\sigma_i^{-2}}{\sum_{j=1}^{n} \sigma_j^{-2}} \tag{7}$$

Other weightings are also possible. A commonly used approach for gridding satellite data is to distribute the satellite pixel based on its overlap with the output grid ("area weighting"). For a given grid box, incorporating the fractional area overlap of the ith satellite pixel with the grid box (F_i) into the weighting yields the following expression.

$$w_i = \frac{F_i \sigma_i^{-2}}{\sum_{j=1}^{n} F_j \sigma_j^{-2}} \tag{8}$$

The sample variance for the grid box ($\sigma_{\overline{\Omega}}^2$) is

$$\sigma_{\overline{\Omega}}^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 \tag{9}$$

Note that the σ_i values are estimated from spectral fit residuals, and thus $\sigma_{\overline{\Omega}}^2$ only reflects the uncertainties due to instrument random noise. It does not account for systematic uncertainties, such as errors in reference cross sections. It is nevertheless useful, as it provides a lower bound on the grid box uncertainties, and can help determine appropriate spatial resolutions and averaging periods for gridding.

References

C. Chan Miller, G. Gonzalez Abad, H. Wang, X. Liu, T. Kurosu, D. J. Jacob, and K. Chance. Glyoxal retrieval from the ozone monitoring instrument. *Atmospheric Measurement Techniques*, 7(11):3891–3907, 2014. doi: 10.5194/amt-7-3891-2014. URL http://www.atmos-meas-tech.net/7/3891/2014/.

Kelly Chance, Thomas P. Kurosu, and Christopher E. Sioris. Undersampling correction for array detector-based satellite spectrometers. *Appl. Opt.*, 44(7):1296–1304, Mar 2005.

Q. L. Kleipool, M. R. Dobber, J. F. de Haan, and P. F. Levelt. Earth surface reflectance climatology from 3 years of omi data. *Journal of Geophysical Research: Atmospheres*, 113(D18):n/a–n/a, 2008. ISSN 2156-2202.